

Valencia, 28 de noviembre de 2011

Investigadores de la UPV y la UJI presentan en el SuperComputing en Seattle una tecnología que permite reducir costes en clusters de altas prestaciones

- rCUDA, la tecnología desarrollada por investigadores de la Universitat Politècnica de València y la Universitat Jaume I de Castellón, fue una de las grandes protagonistas en la exposición internacional de mayor prestigio en el campo de la supercomputación
- Con rCUDA es posible que los centenares o miles de nodos que conforman un cluster de ordenadores de altas prestaciones compartan los acelerados gráficos instalados en él, consiguiendo el 100% de eficiencia

rCUDA, una nueva tecnología desarrollada por investigadores de la Universitat Politècnica de València y la Universitat Jaume I de Castellón, fue una de las protagonistas en la exposición internacional de mayor prestigio en el campo de la supercomputación, SuperComputing 2011, celebrada en Seattle la semana pasada. La herramienta diseñada desde la UJI y la UPV permite el acceso remoto a aceleradores gráficos (GPUs) en un cluster de computadores de altas prestaciones (clúster HPC).

El sistema planteado consiste en una capa software que permite que un programa que se está ejecutando en uno de los computadores del cluster use, a través de la red, los aceleradores gráficos instalados en el cluster, independientemente del computador concreto al que estén conectados, con el fin de acelerar operaciones complejas de cálculo. Sin la tecnología rCUDA, un programa que se está ejecutando en uno de los computadores solo puede usar el acelerador instalado en su mismo nodo. Con rCUDA es posible que los centenares o miles de nodos que conforman un cluster de altas prestaciones compartan los acelerados gráficos instalados en él, consiguiendo el 100% de eficiencia. Esto permite ahorrar energía, al poderse utilizar un menor número de procesadores gráficos, dado que ya no hay que instalar un acelerador en cada nodo del cluster, reduciendo también la inversión en material y disminuyendo el gasto en mantenimiento.

“Esta tecnología supone un paso muy importante para optimizar el rendimiento de los centros de datos y centros de computación de altas prestaciones que utilizan GPUs para sacar más rendimiento a sus aplicaciones, reduciendo el tiempo de ejecución. Imaginemos que hay que hacer una simulación de un volcán; si se hace sin tarjeta gráfica el proceso puede tardar días, dado que tendría que ejecutarse completamente en la CPU del computador, que aunque son muy potentes hoy en día, tienen sus limitaciones; el uso de una GPU lo acelera notablemente, dado que las GPUs realizan las operaciones de cómputo de manera más eficiente que las CPUs. Si empleamos rCUDA lograremos además una reducción significativa de consumo energético”, apunta Federico Silla, investigador del Grupo de Arquitecturas Paralelas (GAP) de la UPV.

La razón para esta reducción en el consumo es que sin rCUDA habría que instalar una GPU en cada uno de los nodos del cluster. “No obstante, lo habitual en este caso es que las GPUs no se usen el 100% del tiempo. Sin embargo, cuando no están realizando cálculos, las GPUs siguen consumiendo una cantidad nada despreciable de energía, que termina siendo desperdiciada. Con rCUDA podemos reducir el número de GPUs en el cluster, y por tanto consumir menos energía. Además, al compartir las GPUs entre diferentes aplicaciones que se están ejecutando al mismo tiempo en los diversos nodos del cluster, aumentamos su utilización, haciendo un uso mucho más eficiente de la energía consumida”, apunta Federico Silla, investigador del Grupo de Arquitecturas Paralelas (GAP) de la UPV.

La aplicación ha sido desarrollada por el estudiante de doctorado de la UPV, Antonio J. Peña, dentro de la investigación que realiza actualmente para su tesis doctoral en la Universitat Jaume I, dirigida por los



profesores Federico Silla – GAP UPV – y Rafael Mayo, del Grupo de Arquitecturas y Computación de Altas Prestaciones de la Universitat Jaume I, quienes también participaron en la demostración.

La demostración realizada en la feria de supercomputación fue posible gracias a la invitación de la empresa Mellanox Technologies, la compañía más importante en la implementación de la tecnología InfiniBand, un protocolo de comunicaciones abierto a cualquier empresa y un proveedor de servicios para servidores y almacenamiento. “Mellanox se interesó por nuestra tecnología ya en el Supercomputing de 2010 en Nueva Orleans. Mellanox considera que nuestra tecnología es altamente innovadora y permite utilizar de forma muy flexible los aceleradores basados en GPUs, siendo un excelente complemento a sus productos. Además, nos ha invitado a presentar nuestro trabajo en diversas conferencias y foros en Suiza, Alemania, y China”, destaca Federico Silla.

Antonio Peña, que actualmente se encuentra en el Centro de Supercomputación de Suiza probando este sistema, defenderá próximamente su trabajo de investigación. Según el profesor Mayo, la aplicación práctica de esta tesis “es el ciclo perfecto para un trabajo doctoral: hacer una investigación que después se pueda aplicar a sistemas reales mejorando su eficiencia”.

Los grupos de investigación de Arquitecturas Paralelas de la Universitat Politècnica de València y de Arquitecturas y Computación de Altas Prestaciones de la Universitat Jaume I trabajan conjuntamente en esta tecnología desde hace más de dos años. El equipo de trabajo de rCUDA está formado, además del doctorando y sus directores de tesis, por José Duato, Javier Nadal y Carlos Reaño, de la UPV y Enrique Quintana y Adrián Castelló de la UJI.

El mundo de la computación trabaja para conseguir en 2020 el primer computador con capacidad *exaflop* que permitirá hacer 10^{18} operaciones por segundo y resolver cálculos de cuestiones relacionadas con el cambio climático o la secuenciación genómica a las que la actual tecnología no puede dar respuesta; y entre las tecnologías que se utilizan para desarrollar este nuevo computador se encuentra el GPU Computing (el uso de la unidad de procesamiento gráfico para realizar operaciones de cálculo científico).

Datos de contacto: Luis Zurano Conches

Unidad de Comunicación Científica-CTT
Universitat Politècnica de València
cienciaupv@upv.es
647422347

Anexos: